THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Early signals of vaccine driven perturbation seen in pneumococcal carriage population genomic data

Citation for published version:
Chaguza, C, Heinsbroek, E, Gladstone, RA, Tafatatha, T, Alaerts, M, Peno, C, Cornick, JE, Musicha, P, Bar-Zeev, N, Kamng'ona, A, Kadioglu, A, McGee, L, Hanage, WP, Breiman, RF, Heyderman, RS, French, N, Everett, DB & Bentley, SD 2019, 'Early signals of vaccine driven perturbation seen in pneumococcal carriage population genomic data', *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciz404

Digital Object Identifier (DOI):
10.1093/cid/ciz404

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Clinical Infectious Diseases

Publisher Rights Statement:
The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 07. May. 2023

Early signals of vaccine driven perturbation seen in pneumococcal carriage population genomic data

Chrispin Chaguza[1,2,3]*, Ellen Heinsbroek[2,4], Rebecca A. Gladstone[1], Terence Tafatatha[5], Maaike Alaerts[3,6], Chikondi Peno[3,7], Jennifer E. Cornick[2,3], Patrick Musicha[2,3,8,9], Naor Bar-Zeev[2,3,10], Arox Kamng'ona[2,3,11], Aras Kadioglu[2], Lesley McGee[12], William P. Hanage[13], Robert F. Breiman[14], Robert S. Heyderman[3,15], Neil French[2,3,†], Dean B. Everett[3,7,†] and Stephen D. Bentley[1,2,16,†]

[1]Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

[2]Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

[3]Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi

[4]HIV & STI Department, National Infection Service, Public Health England, London, UK

[5]Malawi Epidemiology Intervention Research Unit (formerly KPS), Chilumba, Malawi

[6]Center of Medical Genetics, University of Antwerp, Antwerp, Belgium

[7]MRC Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, UK

[8]Mahidol Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand

[9]Nuffield Department of Medicine, University of Oxford, Oxford, UK

[10]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

[11]Department of Biomedical Sciences, University of Malawi, College of Medicine, Blantyre, Malawi

[12]Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, USA

[13]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[14]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, USA

[15]Division of Infection and Immunity, University College London, London, UK

[16]Department of Pathology, University of Cambridge, Cambridge, UK

*Corresponding author: Chrispin Chaguza PhD, Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, CB10 1SA, UK (cc19@sanger.ac.uk).

†N.F., D.B.E. and S.D.B. contributed equally to this work.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Summary

Frequency of carried vaccine serotypes decreased significantly across lineages post-PCV13 introduction in Malawi. There is an increase of serotypes 7C, 15B/C, 23A and 28F post-PCV13 introduction. Antibiotic resistance rates remained stable but certain accessory genes changed in frequency post-PCV.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Abstract

Background. Pneumococcal conjugate vaccines (PCV) have reduced pneumococcal diseases globally. Pneumococcal genomic surveys elucidate PCV effects on population structure but are rarely conducted in low-income settings despite the high disease burden.

Methods. We undertook whole genome sequencing of 660 pneumococcal isolates collected through surveys from healthy carriers two years from PCV14 introduction and one-year post-rollout in northern Malawi. We investigated changes in population structure, within-lineage serotype dynamics, serotype diversity, and frequency of antibiotic resistance (ABR) and accessory genes.

Results. In the under-fives, frequency and diversity of vaccine serotypes (VT) decreased significantly post-PCV but no significant changes occurred in over-fives. Clearance of VT serotypes was consistent across different genetic backgrounds (lineages). There was an increase of non-vaccine serotypes (NVT) namely 7C, 15B/C, 23A in under-fives but 28F increased in both age groups. While carriage rates have been recently shown to remain stable post-PCV due replacement serotypes, there was no change in diversity of NVTs. Additionally, frequency of intermediate-penicillin-resistant lineages decreased post-PCV. While frequency of ABR genes remained stable, other accessory genes especially those associated with MGEs and bacteriocins showed changes in frequency post-PCV.

Conclusions. We demonstrate evidence of significant population restructuring post-PCV driven by decreasing frequency of vaccine serotypes and increasing frequency of few NVTs mainly in under-fives. Continued surveillance with WGS remains crucial to fully understand dynamics of the residual VTs and replacement NVT serotypes post-PCV.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Keywords. Pneumococcus, carriage, genomics, serotypes

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Introduction

Pneumococcal conjugate vaccines (PCV) have demonstrated high effectiveness on non-invasive [1] and invasive pneumococcal disease (IPD) [2]. Approximately 70% reduction of IPD was observed in USA post-PCV7 introduction with >90% reduction of vaccine serotypes (VTs) [2]. PCV7 was not widely introduced in Sub Saharan Africa (SSA) due to predicted low VT coverage against common serotypes including 1 and 5 [3]. South Africa was among few countries in SSA to implement PCV7 and >85% reduction in IPD which included HIV-infected individuals was reported [4]. Subsequent higher-valent PCVs (PCV10 and PCV13) have shown high effectiveness in SSA in on VT carriage (>65%) [5] and IPD (>80%) in SSA [6, 7] consistent with reports in high-income countries [8].

While reduction of VT carriage has been documented post-PCV [9], the effect of PCV on overall carriage rates and density is not substantial [10, 11] because of serotype replacement post-PCV (increase of NVTs) [12, 13]. Such replacement is prompted by rewiring of strain competition dynamics when VTs become uncommon. The replacement NVTs can be novel (imported or capsule-switched) or extant due to clonal expansion of previously suppressed NVT serotypes. While replacement NVTs are less invasive than VTs, this is not universally true, and some strains retain propensity for IPD [14]. Therefore, post-PCV surveillance is crucial to monitor serotype frequency and interactions to inform optimal future PCV formulations. Pneumococcal surveillance typically focuses on IPD but monitoring of carriage is also crucial because that's the niche where evolution occurs therefore influence population-level dynamics unlike IPD which is an evolutionary dead-end.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

PCV13 was introduced in Malawi on 12 November 2011 as an accelerated '3+0' schedule at 6, 10 and 14 weeks with a limited catch-up for infants in the first year from introduction. For the catch-up campaign, infants aged <12 months at the date of vaccine introduction were eligible for the following 12 months to receive 3 doses of PCV13 but coverage for the three doses of PCV13 was lower in catch-up infants (~50-70%) than in birth-eligible children (~90-95%) [15]. Recent carriage studies have reported a reduction of VTs but no overall change in carriage rates [16]. In this study, we undertook whole genome sequencing (WGS) of pneumococcal isolates from northern Malawi to investigate early changes in population re-structuring, within-lineage serotype composition, serotype diversity, antibiotic resistance (ABR) and accessory genome dynamics pre- and post-PCV. The WGS was part of the Global Pneumococcal Sequencing (GPS) project (www.pneumogen.net), which has sequenced ~23,000 isolates globally to study pneumococcal evolution patterns post-PCV to inform future vaccine design.

Materials and methods

Study population and isolate selection

Household surveys of healthy carriers were conducted in Karonga district of northern Malawi pre-PCV between 2009-2011 and post-PCV in 2014 and there was no significant changes of the overall pneumococcal carriage rates [16]. We randomly selected a subset of the samples (n=660) pre-PCV in 2009-2010 (n=482) and post-PCV in 2014 (n=178) for a WGS survey (Supplementary Data 1). The mean age for samples collected pre-PCV was 6.24 (95% CI: 5.30-7.17) and 4.84 (95% CI: 3.79-5.89) post-PCV. The age ranged from 3 days to 54 years old and one month to 30 years old pre-PCV and post-PCV respectively. In terms of age group, 506 samples collected pre-PCV were from under-fives (n=376, mean: 1.70 years old [95% CI: 1.53 to 1.86]) and over-fives (n=130, mean: 22.33 [95% CI: 19.91-24.75]) while 154 samples were

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

collected post-PCV in under-fives (n=106, mean: 1.35 [95% CI: 1.05-1.65]) and over-fives (n=48, mean: 14.29 [95% CI: 12.08-16.50]).

The nasopharyngeal swabs were processed as previously described [16]. Informed written consent was obtained from adults, and parents, guardians and caregivers of child participants. The study was approved by National Health Sciences Research Committee in Malawi (#490 and #1232), London School of Hygiene and Tropical Medicine (#5345), University of Liverpool (#670) and University of Malawi College of Medicine Research Ethics Committee (#P.O8/14/1614).

Genomic DNA sequencing and analysis

Procedures for genomic DNA extraction and sequencing were described previously [17]. The sequence reads were assembled using an automated assembly pipeline [18]. The serotypes and sequence types (ST) were identified using seroBA [19] and multi-locus sequence typing [20] while ABR genes were detected using nucleotide-BLAST v2.2.30 (E-value<0.001, sequence coverage and identity >80%) [21]. Penicillin minimum inhibitory concentrations (MIC) were genotypically predicted using the Centers for Disease Control and Prevention's (CDC) pipeline [22] and the MICs were interpreted using the British Society for Antimicrobial Chemotherapy breakpoints [23]. The sequence reads were deposited in the European Nucleotide Archive under accession numbers in Supplementary Data 1.

Genomic clusters (GC) or lineages were inferred using BAPS v6.0 [24]. A 1,050,021bp core-gene-alignment with 88,961 single nucleotide polymorphism (SNP) positions was generated using Roary [25]. Phylogenetic trees for all isolates was constructed using the core-gene-alignment using FastTree-SSE3 v2.1.3 [26] while lineage-based trees used

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

consensus alignments from SMALT (https://sourceforge.net/projects/smalt/) for Gubbins v1.4.10 [27] and RAxML v7.0.4 [28]. The SNPs were reconstructed on the trees using parsimony. The trees were visualised using iToL v2.1 [29] and MicroReact [30].

Statistical analysis

Serotype and ABR gene frequencies were compared using Fisher's Exact test. The Simpson diversity index (D) for serotype and ST composition were estimated using a web-based analysis tool (www.comparingpartitions.info). The differences in penicillin MICs were assessed using Student's t-test. The binary presence-absence of accessory genes pre- and post-PCV was assessed using logistic regression, which controlled for vaccine status and age group of the isolates with Bonferroni correction for multiple testing. We used R v3.1.2 for statistical analyses (R Core Team, 2013, www.r-project.org).

Results

Defining the population structure

Whole genomes sequences for the 660 carried isolates revealed 45 serotypes and 169 sequence types (STs) (Supplementary Data 1). The majority of the serotypes were associated with a single phylogenetic tree branch with few serotypes dispersed across multiple unrelated branches due to capsule-switching (acquisition of the capsule/serotype in unrelated strains). To account for genetic background of the isolates, we defined 23 lineages or genomic clusters (GC) using unsupervised nucleotide sequence clustering and these lineages allowed for subsequent comparison of serotype changes in context of their lineage membership (Figure 1a). Phylogenetically, all the lineages (GC1-22) except for GC23 were monophyletic with their members emerging from a single recent common ancestor. The defined lineages varied in composition of serotypes and sequence diversity reflecting differences in either evolution rates (higher rate imply

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

higher diversity) or age (older lineages accrue more diversity). An interactive phylogenetic tree of the isolates is available in MicroReact (https://microreact.org/project/xH7-VcoWj/8a339d57).

Frequency of vaccine serotypes and their associated lineages

The frequency of lineages GC3 (P=0.002), GC16 (P=0.004) and GC17 (P=0.004) in under-fives increased post-PCV while lineages GC10 (P=0.016) and GC19 (P=0.026) showed a decrease in the same age group (Figure 2a, Supplementary Table 1). These were predominantly VT-associated lineages therefore their decreased frequency reflects reduction of VTs in isolates sampled from the under-fives, which comprised of the majority of vaccinated individuals, but this reduction was not seen in the unvaccinated over-fives population (Figure 2b, Supplementary Table 2). However, despite the observed decrease in frequency of VTs in lineages, the odds ratio for VT serotypes in each lineage in under-fives relative to over-fives remained unchanged post-PCV except for lineages GC4 and GC16 in which the odds ratio (OR) for VTs increased and lineages GC19 and GC21 where the OR decreased (Supplementary Table 3). By individual serotypes, frequency of multiple NVTs increased post-PCV namely serotypes 7C (P=0.001), 15B/C (P=0.004), 23A (P=0.017) and 28F (P=0.0001) in under-fives and 28F (P=0.029) in over-fives (Figure 2c and Supplementary Table 4). There overall frequency of VTs regardless of lineages decreased post-PCV in under-fives (60.08-33.08%; $P=4.80\times10^{-8}$) but not over-fives (42.45-33.33%; P=0.3739) (Figure 2d and Supplementary Table 5). However, the frequency of VTs was higher in under-fives than over-fives ($P=8.44\times10^{-4}$) pre-PCV but no differences were observed post-PCV (Figure 2d, Supplementary Table 4,5).

Emergence and clonal expansion of NVT isolates

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Within-lineage dynamics revealed how serotype frequency were changing post-PCV. For example, clonal expansion of serotypes 11A (GC7) and 15B/C (GC16) occurred post-PCV (Figure 3). By comparing pre- and post-PCV frequencies of serotypes in the lineages it was clear that the majority of the serotypes were extant pre-PCV while a few were first detected post-PCV in some lineages for example serotype 19F (GC13) and 19A (GC20), which were capsule-switch events. Comparison of overall serotype frequencies revealed that serotype 28F (GC2 and GC21) was the only serotype not detected pre-PCV, which posed questions regarding whether it emerged by recent importation or post-PCV unmasking after prior circulation pre-PCV at undetectable level. Genomic analysis showed that GC21 contained serotype 9V and the 28F amongst 9V strains in GC21 emerged post-PCV by a 9V→28F vaccine capsule-switch and the capsule-switched isolates were distinguished by ~20 SNPs within themselves and with closest 9V isolates. However, accrued genetic diversity of 28F isolates in GC2 was higher than expected to occur under pneumococcal mutation rate if the isolates emerged post-PCV (maximum 6,757 SNPs) (Figure 4). Because recently imported isolates typically undergoes a loss in genetic diversity (bottleneck), our findings were not consistent with this therefore implied that serotype 28F isolates existed at undetectable levels pre-PCV and were unmasked due to clearance of VTs post-PCV, which subsequently led to the serotype 9V→28F vaccine capsule-switch in GC21. There was further evidence against importation from other countries because serotype 28F with similar genetic profiles were uncommon globally (2/12000 in GPS collection) and these differed from our isolates by >2000 SNPs therefore could not be the source. Additional capsule-switched isolates detected post-PCV only included 11A→20 (GC7), 13→19A (GC9), 16F→19F (GC13), (GC21), and 7C→NT (GC23) but none of these were vaccine-escape events.

Serotype diversity as an indicator for PCV effectiveness

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

The pre-PCV equilibrium diversity of serotypes is altered by PCV therefore changes in this diversity in VTs and NVTs quantified using Simpson diversity index (D) can signal population-level PCV effectiveness. In this study, Simpson's D for serotypes decreased post-PCV in VTs in under-fives (P=0.022) but not over-fives (Figure 4a, 6 Table). However, Simpson's D appeared to decrease and increase in NVTs in under-fives and over-fives respectively although not significant statistically (Figure 5). The Simpson's D for serotypes was similar between VTs and NVTs pre-PCV, but it was higher in NVTs than VTs post-PCV (P=0.004) (Figure 5b). The Simpson's D was higher for STs than serotypes both pre-PCV (P=0.011) and post-PCV (P=0) but remain unchanged post-PCV therefore ST diversity may not be informative of PCV effect (Supplementary Table 6,7).

Frequency of antibiotic resistance and other accessory genes

An important subset of pneumococcal accessory genes encodes for ABR-conferring proteins therefore we assessed their distribution pre- and post-PCV. The genes confer resistance against tetracycline, chloramphenicol and erythromycin. No significant changes in frequency of ABR genes (rates) occurred post-PCV (Figure 6a-b, Supplementary Table 8). We also assessed MICs changes in penicillin resistance genotypically and similarly resistance rates were unchanged although the MICs decreased post-PCV (P=0.0098) due to clearance of intermediate-resistant VT isolates in GC12 (Figure 6c).

To assess whether other accessory genes had changed in frequency post-PCV unlike ABR genes, logistic regression model was fitted to binary gene presence-absence data with sampling period of the isolates as independent variable adjusted for PCV status and age group. Significant changes in frequency were detected in forty-two accessory genes post-PCV after Bonferroni correction (Figure 7, Supplementary Table 9). There was an increasing trend in half of the genes with lowest P-values associated with glycosyl transferase (OR=3.34, $P=9.71\times10^{-6}$), bacteriolysin (OR=3.34, $P=3.46\times10^{-5}$) and

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

restriction modification system (OR=3.34, P=3.46×10^{-5}). Conversely, bacteriocin gene *blpQ* (OR=0.19, P=5.20×10^{-6}) showed the most significant decrease. The capsule biosynthesis gene *wzx* also decreased post-PCV (OR=0.43, P=0.042). By functional classification, majority of the detected genes were associated with mobile DNA (11/42) and bacteriocins (3/42).

Discussion

Pneumococcal evolution occurs during carriage therefore monitoring carriage population is crucial to understand ongoing strain dynamics and adaptations post-PCV [4, 5]. Our genomic study demonstrates the utility of WGS in monitoring changes in pneumococcal carriage in low-income settings post-PCV introduction. We provide evidence of early pneumococcal population restructuring using WGS observed beyond serotype but also at lineage and accessory genome level in both children and adults two years post-PCV introduction. The most significant finding is that decrease of VTs occurred consistently across their associated lineages reflecting no interference of genetic background on PCV effectiveness, and most importantly the collective decrease of VTs was more pronounced in under-fives, which implies high direct PCV effects consistent with data from clinical trials [16]. However, our findings show modest decrease in frequency of VTs and VT-associated lineages in over-fives which implies either limited or delayed indirect PCV effects in contrast to other settings where higher indirect effects have been demonstrated in older population [9, 31]. It is currently unclear whether limited indirect PCV effects in our setting is related to PCV scheduling or mitigating factors for herd effects such as HIV-infected adults considered as potential reservoir for pneumococcal diversity post-PCV [32].

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

PCV implementation changes both frequency of individual serotypes and their composition in the population therefore quantifying the degree of disorder of serotypes with Simpson diversity index can indirectly inform population-level PCV effects [33, 34]. Our findings showed that serotype diversity in VTs decreased significantly post-PCV but only in under-fives, which implies substantial loss of diversity in VTs in vaccinated under-fives but not older unvaccinated individuals consistent with the serotype-frequency-based observation of high direct PCV effect but limited indirect effects. We expected an increase in serotype diversity in NVTs due to serotype replacement, on the contrary, this did not occur in both age groups, which implies that while the overall carriage rate was rapidly restored by replacement NVTs, these resulted in marginal accrual of additional diversity in NVTs post-PCV. This may be a consequence of incomplete serotype replacement process because the isolates were sampled only two-years post-PCV introduction. However, increase in frequency of individual NVTs occurred for serotypes 7C, 23A and 15B/C in under-fives, and 28F in both age groups and these have been reported elsewhere except for serotype 28F [35, 36]. Unlike other serotypes, 28F was detected only post-PCV in lineages (GC2 and GC21), which initially prompted speculation that it was imported from another country. However, within-lineage genetic diversity of 28F isolates and genetic dissimilarity from similar isolates in the GPS collection ruled out importation because the genetic diversity was much higher than can be expected for a newly introducted clone, which suggested that it had been circulation pre-PCV but at undetectable levels. Interestingly, the serotype 28F in GC21 emerged post-PCV via a vaccine-escape 9V→28F capsule switch between unmasked 28F in GC2 and 9V isolates in GC21 but there were no other vaccine-escape capsule switches detected for other serotypes. Therefore, the rarity of vaccine-escape capsule-switches demonstrates negligible effect of capsule-switching process on serotype replacement unlike clonal expansion of previously masked capsule-intact NVT isolates.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

The pneumococcal accessory gene pool consists of a diverse repertoire of genes, which include ABR, MGE and competition-associated genes. We noted stability in frequency of ABR genes post-PCV but considering that the levels were already low unlike in IPD isolates in Malawi this did not raise concerns [37-39]. Interestingly, PCV reduced frequency of intermediate-penicillin-resistant isolates associated with VT-lineages, which showcases how PCV can be strategically harnessed as a preventative strategy to thwart emergence of high resistance by targeting low-level resistance lineages with highest likelihood to express full resistance in future [40]. Other non-ABR-associated accessory genes showed changes in frequency post-PCV after controlling for age group and vaccine status of the isolates and these predominantly included highly mobile and rapidly shared MGE-associated genes and bacteriocin immunity proteins, which mediates isolate competition therefore suggests potential benefits in emerging NVTs as pneumococcal population re-establishes equilibrium serotype dynamics.

Our findings demonstrate early changes in the pneumococcal carriage population in a low-income setting post-PCV introduction. Our findings provide the first large-scale post-PCV genomic survey of carried pneumococcal isolates in an African setting where despite high disease burden, limited WGS studies are conducted therefore this study provides useful baseline data for comparative analyses of population-level effects of PCV between different settings both low and high-income. The play of chance and possibly small sample sizes for subset analyses could have impacted some of our findings therefore cautious interpretation is recommended. Continued diligent surveillance and WGS remain crucial for monitoring long-term residual effects VTs, serotype replacement and genotypic changes post-PCV after equilibrium serotype dynamics are re-stablished. Additionally, our study complements vaccine efficacy data from clinical trials therefore improves our understanding of population-level effects of PCV in Malawi,

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

SSA and globally, which will help to inform optimal combination of serotypes for future PCVs to maximise their beneficial effects especially in vulnerable tropical populations.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Author contributions

CC, NF, DBE and SDB conceived and designed the study. NF, DBE and SDB supervised the study. EH, TT and NF conducted the field studies. MA, AWK, JEC and CP performed molecular and microbiology experiments. SDB supervised whole genome sequencing and genomic analysis. CC and RAG checked quality of the sequence assemblies. CC performed genomic and statistical analyses. CC and SDB wrote initial draft of the paper. PM, LM, RFB, AK, WPH and RSH contributed to data interpretation. All authors contributed to writing and reviewing of the paper.

Acknowledgements

We acknowledge work by clinical and laboratory staff at Malawi Epidemiology and Intervention Research Unit (MEIRU) and Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW), and sequencing and informatics teams at Wellcome Sanger Institute.

Financial support

This work was supported by Bill and Melinda Gates Foundation for funding the Global Pneumococcal Sequencing (GPS) project [www.pneumogen.net] (grant number: OPP1034556 to SDB, LM and RFB) and Wellcome UK (core grant number: 084679/Z/08/Z to MLW). RSH and NF are supported by UK Medical Research Council (MRC) and the UK Department for International Development (DFID) grant under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union (MR/N023129/1). CC was funded by a PhD studentship from the Commonwealth Scholarship Commission in the UK.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Potential conflicts of interest

NBZ reports investigator-initiated project grants from GlaxoSmithKline Biologicals and from Takeda Pharmaceuticals, outside the submitted work; RAG reports PhD studentship from Pfizer 2009-2013, outside the submitted work; and WPH reports personal fees from Antigen Discovery Inc., outside the submitted work. Other authors declare no conflict of interests.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

References

1. Ben-Shimol S, Givon-Lavi N, Leibovitz E, Raiz S, Greenberg D, Dagan R. Impact of Widespread Introduction of Pneumococcal Conjugate Vaccines on Pneumococcal and Nonpneumococcal Otitis Media. Clin Infect Dis **2016**; 63(5): 611-8.

2. Whitney CG, Farley MM, Hadler J, et al. Decline in Invasive Pneumococcal Disease after the Introduction of Protein-Polysaccharide Conjugate Vaccine. New England Journal of Medicine **2003**; 348(18): 1737-46.

3. Gordon SB, Kanyanda S, Walsh AL, et al. Poor potential coverage for 7-valent pneumococcal conjugate vaccine, Malawi. Emerging infectious diseases **2003**; 9(6): 747-9.

4. von Gottberg A, de Gouveia L, Tempia S, et al. Effects of vaccination on invasive pneumococcal disease in South Africa. N Engl J Med **2014**; 371(20): 1889-99.

5. Hammitt LL, Akech DO, Morpeth SC, et al. Population effect of 10-valent pneumococcal conjugate vaccine on nasopharyngeal carriage of Streptococcus pneumoniae and non-typeable Haemophilus influenzae in Kilifi, Kenya: findings from cross-sectional carriage studies. Lancet Glob Health **2014**; 2(7): e397-405.

6. Mackenzie GA, Hill PC, Jeffries DJ, et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. The Lancet infectious diseases **2016**; 16(6): 703-11.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

7. Cohen C, von Mollendorf C, de Gouveia L, et al. Effectiveness of the 13-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in South African children: a case-control study. The Lancet Global Health **2017**.

8. Moore MR, Link-Gelles R, Schaffner W, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. The Lancet infectious diseases **2015**.

9. van Hoek AJ, Sheppard CL, Andrews NJ, et al. Pneumococcal carriage in children and adults two years after introduction of the thirteen valent pneumococcal conjugate vaccine in England. Vaccine **2014**; 32(34): 4349-55.

10. Brugger SD, Frey P, Aebi S, Hinds J, Muhlemann K. Multiple colonization with S. pneumoniae before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. PLoS One **2010**; 5(7): e11638.

11. Dunne EM, Manning J, Russell FM, Robins-Browne RM, Mulholland EK, Satzke C. Effect of pneumococcal vaccination on nasopharyngeal carriage of Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis, and Staphylococcus aureus in Fijian children. J Clin Microbiol **2012**; 50(3): 1034-8.

12. Croucher NJ, Finkelstein JA, Pelton SI, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet **2013**; 45(6): 656-63.

13. Gladstone RA, Jefferies JM, Tocheva AS, et al. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. Vaccine **2015**; 33(17): 2015-21.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

14. Assaf R, Shalom B-S, Zinaida K, et al. Emergence of Streptococcus pneumoniae Serotype 12F after Sequential Introduction of 7- and 13-Valent Vaccines, Israel. Emerging Infectious Disease journal **2018**; 24(3): 453.

15. Mvula H, Heinsbroek E, Chihana M, et al. Predictors of Uptake and Timeliness of Newly Introduced Pneumococcal and Rotavirus Vaccines, and of Measles Vaccine in Rural Malawi: A Population Cohort Study. PLoS One **2016**; 11(5): e0154997.

16. Heinsbroek E, Tafatatha T, Phiri A, et al. Pneumococcal carriage in households in Karonga District, Malawi, before and after introduction of 13-valent pneumococcal conjugate vaccination. Vaccine **2018**.

17. Everett DB, Cornick J, Denis B, et al. Genetic characterisation of Malawian pneumococci prior to the roll-out of the PCV13 vaccine using a high-throughput whole genome sequencing approach. PLoS One **2012**; 7.

18. Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microbial Genomics **2016**; 2(8).

19. Epping L, van Tonder AJ, Gladstone RA, et al. SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. Microb Genom **2018**.

20. Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology **1998**; 144 (Pt 11): 3049-60.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

21. Altschul S, Madden T, Schaffer A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **1997**; 25: 3389 - 402.

22. Li Y, Metcalf BJ, Chochua S, et al. Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting β-Lactam Resistance Levels in Streptococcus pneumoniae. mBio **2016**; 7(3).

23. Andrews JM. BSAC standardized disc susceptibility testing method. The Journal of antimicrobial chemotherapy **2001**; 48 Suppl 1: 43-57.

24. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Molecular Biology and Evolution **2013**.

25. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics **2015**; 31(22): 3691-3.

26. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution **2009**; 26(7): 1641-50.

27. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research **2014**.

28. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **2006**; 22: 2688 - 90.

29. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Research **2011**; 39(suppl 2): W475-W8.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

30. Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microbial Genomics **2016**; 2(11).

31. Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. The Lancet infectious diseases **2011**; 11(10): 760-8.

32. Heinsbroek E, Tafatatha T, Phiri A, et al. Persisting high prevalence of pneumococcal carriage among HIV-infected adults receiving antiretroviral therapy in Malawi: a cohort study. Aids **2015**; 29(14): 1837-44.

33. Hanage WP, Finkelstein JA, Huang SS, et al. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. Epidemics **2010**; 2(2): 80-4.

34. Chang Q, Stevenson AE, Croucher NJ, et al. Stability of the pneumococcal population structure in Massachusetts as PCV13 was introduced. BMC infectious diseases **2015**; 15(1): 68.

35. Su L-H, Kuo A-J, Chia J-H, et al. Evolving pneumococcal serotypes and sequence types in relation to high antibiotic stress and conditional pneumococcal immunization. Scientific Reports **2015**; 5: 15843.

36. Ashley M, Shamez NL, Georgia K, Ella C, Norman KF, Carmen S. Rapid Spread of Pneumococcal Nonvaccine Serotype 7C Previously Associated with Vaccine Serotype 19F, England and Wales. Emerging Infectious Disease journal **2018**; 24(10): 1919.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

37. Chaguza C, Cornick JE, Andam CP, et al. Population genetic structure, antibiotic resistance, capsule switching and evolution of invasive pneumococci before conjugate vaccination in Malawi. Vaccine **2017**; 35(35 Pt B): 4594-602.

38. Everett DB, Mukaka M, Denis B, et al. Ten years of surveillance for invasive Streptococcus pneumoniae during the era of antiretroviral scale-up and cotrimoxazole prophylaxis in Malawi. PLoS One **2011**; 6(3): e17765.

39. Musicha P, Cornick JE, Bar-Zeev N, et al. Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998-2016): a surveillance study. The Lancet infectious diseases **2017**; 17(10): 1042-52.

40. Lipsitch M, Siber GR. How Can Vaccines Contribute to Solving the Antimicrobial Resistance Problem? mBio **2016**; 7(3).

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figures

Figure 1. Sampling location, genetic similarity and distribution of carried pneumococcal isolates. The map of Africa shows the location of Malawi and Karonga district where the isolates were sampled from. The number of isolates (n=660) used in the genomic analysis are shown in the table below the phylogenetic tree. The core genome maximum likelihood phylogenetic tree of the 660 carriage isolates rooted at the branch of 'classical' non-typeables (NT). The tips (circles) of the tree are coloured by serotype and coloured strips panels to the right correspond to serotype, genomic clusters (GC) or lineage, vaccine status, sampling period and age group. The tree with metadata and corresponding international definitions of the pneumococcal lineages is available interactively online at https://microreact.org/project/xH7-VcoWj/8a339d57.

Figure 2. Frequency of lineages and serotypes in carriage. (a) Frequency of lineages pre- and post-PCV in under-fives and over-fives. (b) Frequency of VTs in lineages pre- and post-PCV in under-fives and over-fives. (c) Volcano plots showing odds ratio (OR) of individual serotypes pre- and post-PCV in under-fives and over-fives. The x-axis show magnitude ($\log_2$ [OR]) and y-axis show statistical significance ($-\log_{10}$ *P*-value). (d) Frequency of VTs in under-fives and over-fives. Statistically significant changes are marked as follows: 'ns': not significant, $P<0{\cdot}05$ (*), $P<0{\cdot}01$ (**) and $P<0{\cdot}001$ (***). The comparative estimates of prevalence for serotypes, lineages and odds ratios are provided in Supplementary Tables 1-5.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 3. Dynamics of pneumococcal lineages and serotypes. The leftward facing stacked bar graph shows frequency of lineages in under-fives while the rightward facing bar graph shows frequency of lineages and their constituent serotypes in over-fives pre- and post-PCV introduction. The bar graphs are aligned by genomic clusters (GC) for easy comparisons of frequency of serotypes pre- and post-PCV between the two age groups. The serotypes are distinguished by different colours in the bar graphs as described in the key. The GC23 is the 'bin' cluster because it consists of isolates not placed in monophyletic clusters GC1-22. The lineages whose frequency changed significantly post-PCV are marked as follows: $P<0.05$ (*) and $P<0.01$ (**) and those with borderline significance $P<0.095$ (.). The Fisher's exact test was used to determine P-values.

Figure 4. **Genetic diversity of a recently emerged serotypes (28F).** Boxplots showing within (Malawi) and between country (Malawi and South Africa) genetic diversity of serotype 28F isolates showing in (a) GC2 and (b) GC21. Lineage GC21 also include serotype 9V isolates, which some of which underwent a capsule-switch to acquire a serotype 28F capsule.

Figure 4. **Serotype composition and diversity in context of PCV.** (a) Simpson diversity index for composition of serotypes between pre- and post-PCV datasets among VT, NVT and all isolates among isolates. (b) Simpson diversity index for composition of serotypes between VT and NVT isolates sampled pre- and post-PCV. Statistically significant changes are marked as follows: 'ns': not significant, $P<0.05$ (*), $P<0.01$ (**) and $P<0.001$ (***). The estimates and P-values for frequency of VTs and Simpson diversity are provided in Supplementary Table 6,7.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 5. **Genetic diversity of a recently emerged serotypes (28F).** Boxplots showing within (Malawi) and between country (Malawi and South Africa) genetic diversity of serotype 28F isolates showing in (a) GC2 and (b) GC21. Lineage GC21 also include serotype 9V isolates, which some of which underwent a capsule-switch to acquire a serotype 28F capsule.

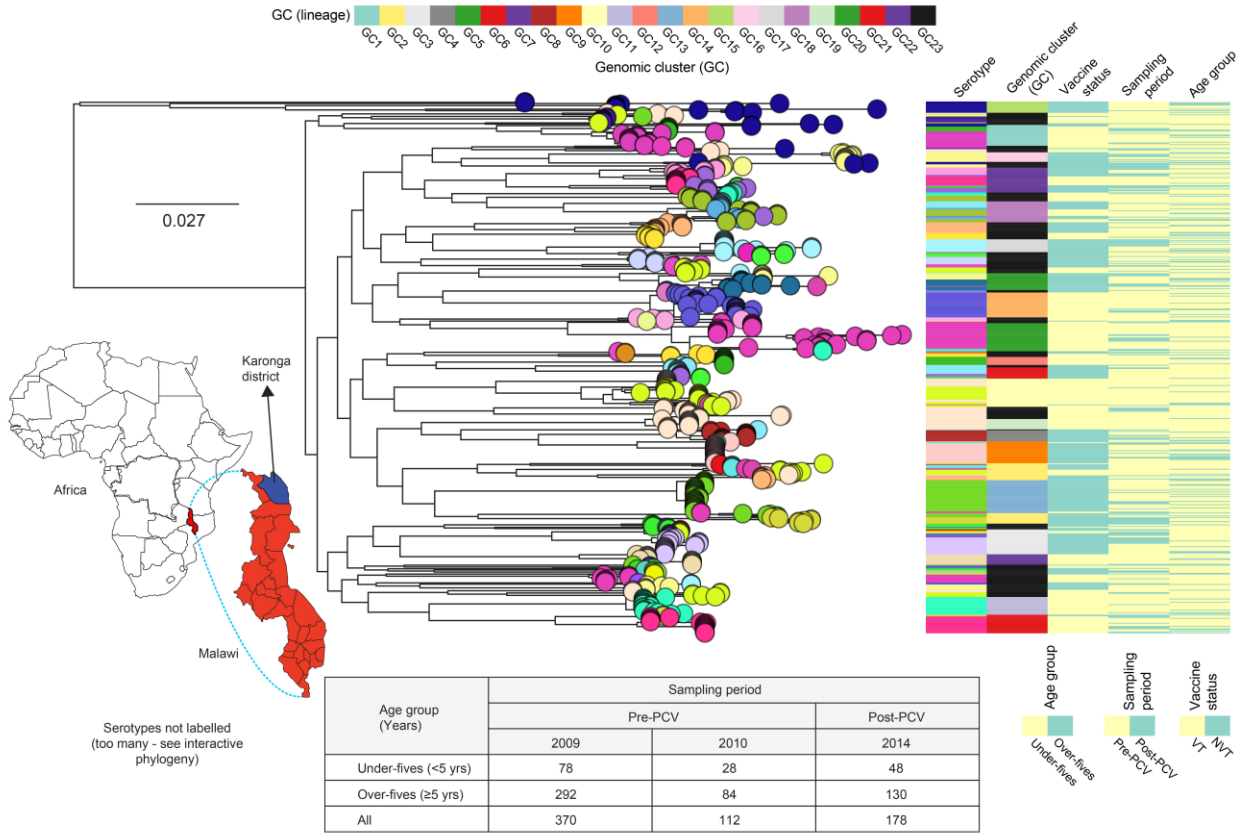
Figure 6. Distribution of antibiotic resistance genes and MGEs pre- and post-PCV. (a) Distribution of chloramphenicol resistance gene (cat_{pC194}), erythromycin resistance gene (*mefA, mefE* and *ermB*), tetracycline resistance gene (*tetM*) and penicillin resistance genes across the phylogenetic tree of the carried pneumococcal isolates. Presence and absence of genes is indicated by colored branches and innermost ring surrounding the phylogeny as shown in the key at the bottom of each tree. (b) Frequency of genotypic ABR rates for chloramphenicol, erythromycin, tetracycline and intermediate-penicillin-resistance pre- and post-PCV. (c) Distribution of penicillin MICs pre- and post-PCV. The subsets with statistically significant changes are marked as follows: 'ns': not significant, $P<0.05$ (*), $P<0.01$ (**) and $P<0.001$ (***). The estimates for frequency of the ABR genes are provided in Supplementary Table 8.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 7. Pneumococcal accessory genome dynamics. The distribution of 2,591 intermediate-frequency-accessory genes in the entire pneumococcal population. The volcano plot shows magnitude (log_2 OR) on the x-axis and statistical significance ($-log_{10}$ P-value) and odds ratio for presence of accessory genes post-PCV relative to pre-PCV after controlling for vaccine status and age group of the isolates. The points were coloured by adjusted P-values after correcting for multiple testing using Bonferroni method. The estimates for the OR and P-values are provided in Supplementary Table 9.

Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

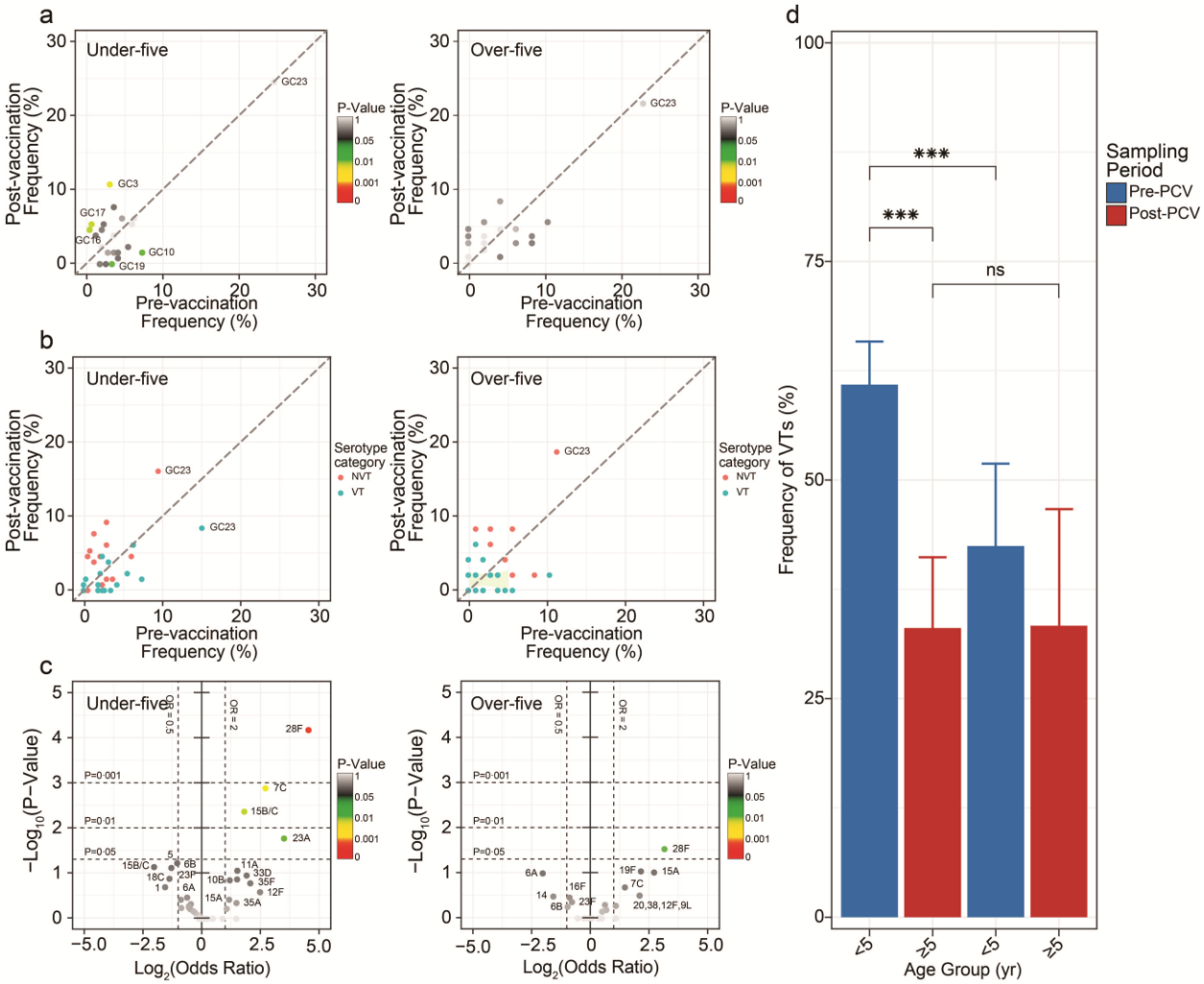
Figure 1



Age group (Years)	Sampling period		
	Pre-PCV		Post-PCV
	2009	2010	2014
Under-fives (<5 yrs)	78	28	48
Over-fives (≥5 yrs)	292	84	130
All	370	112	178

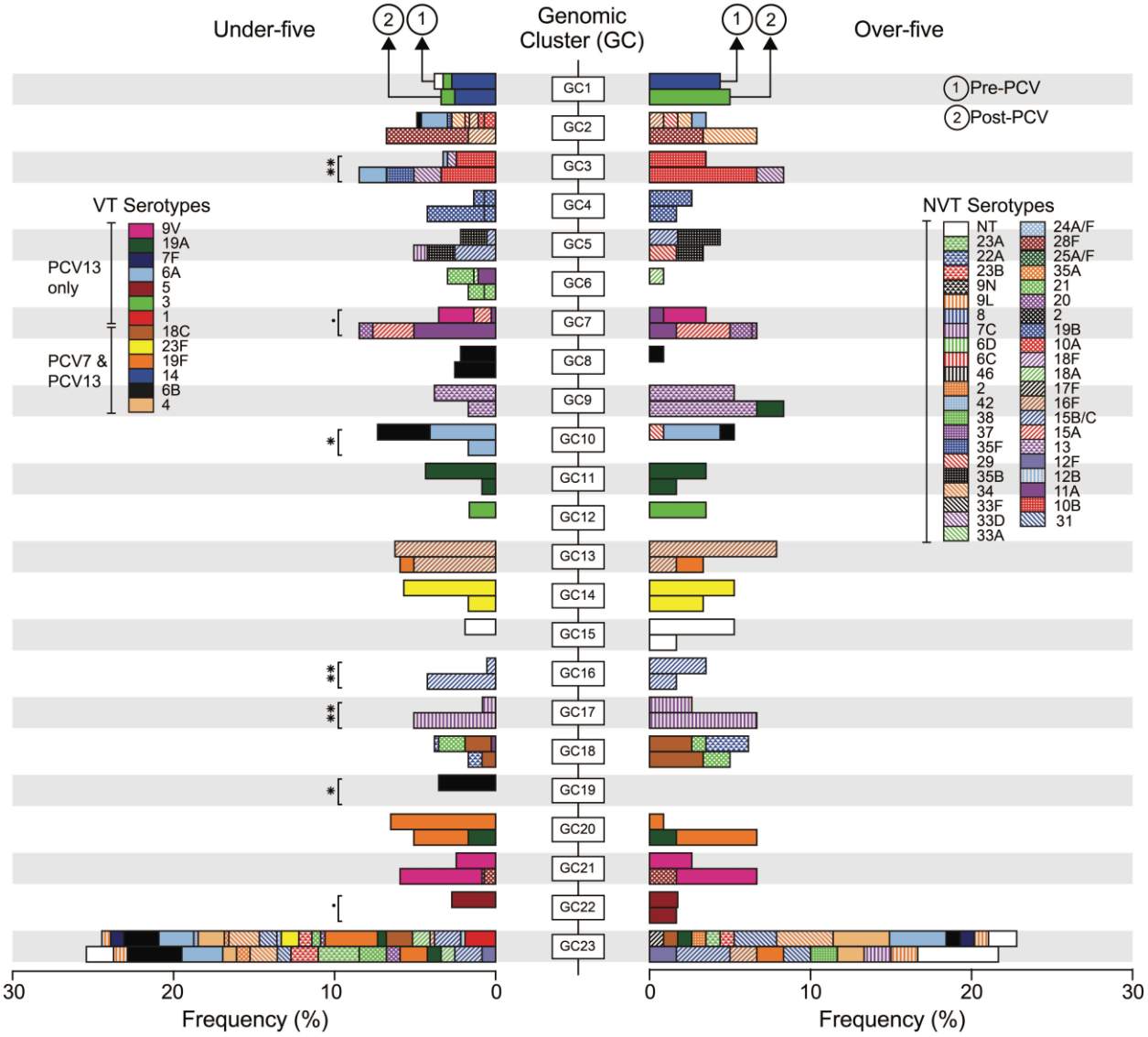
Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 2



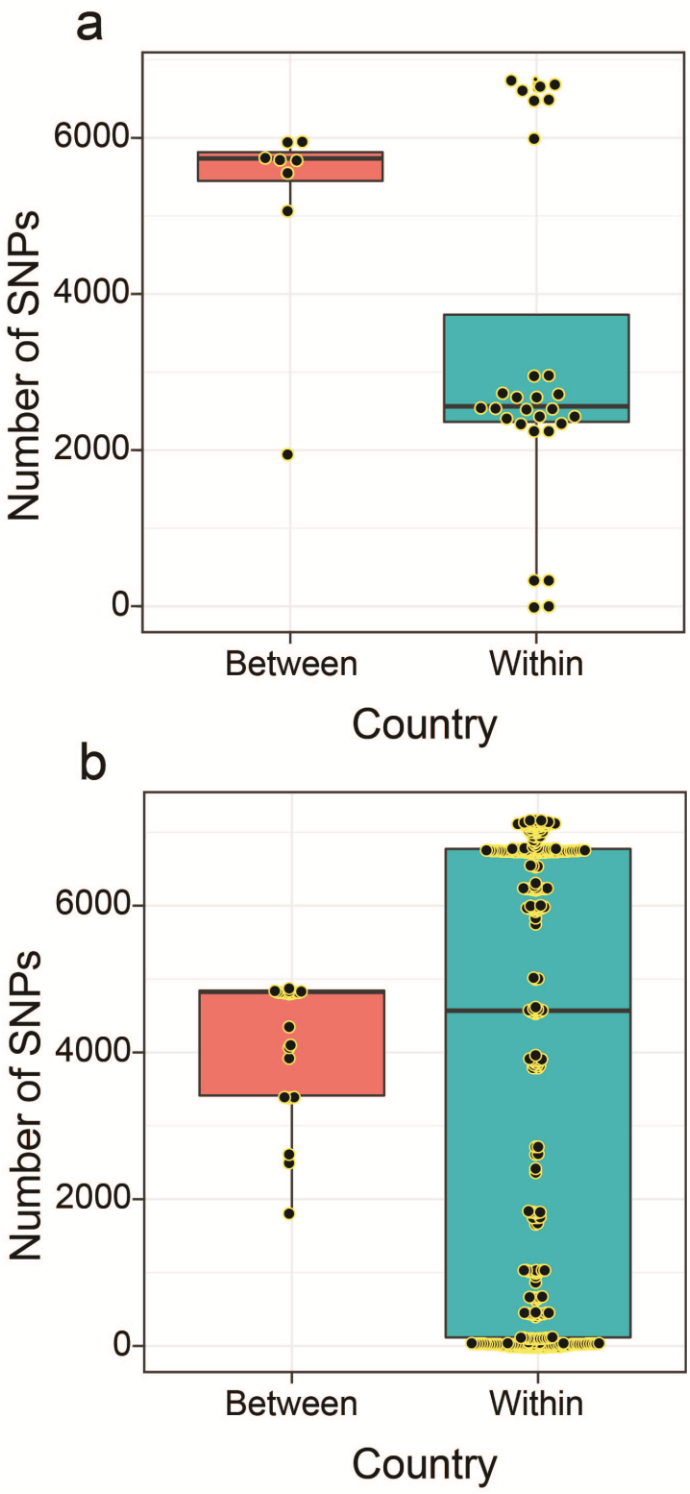
Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 3



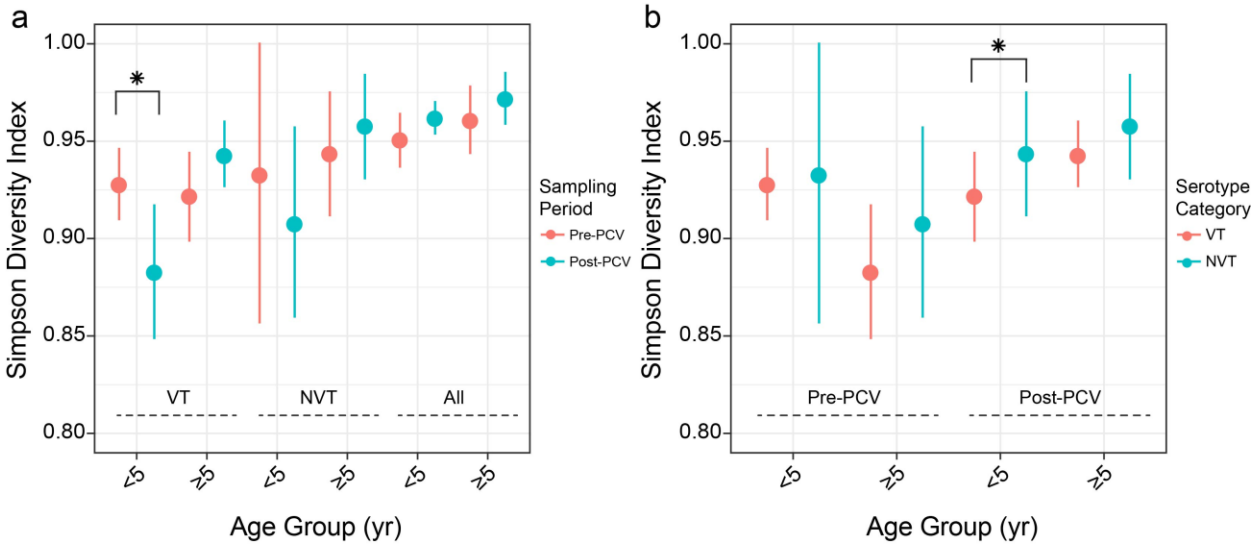
Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 4



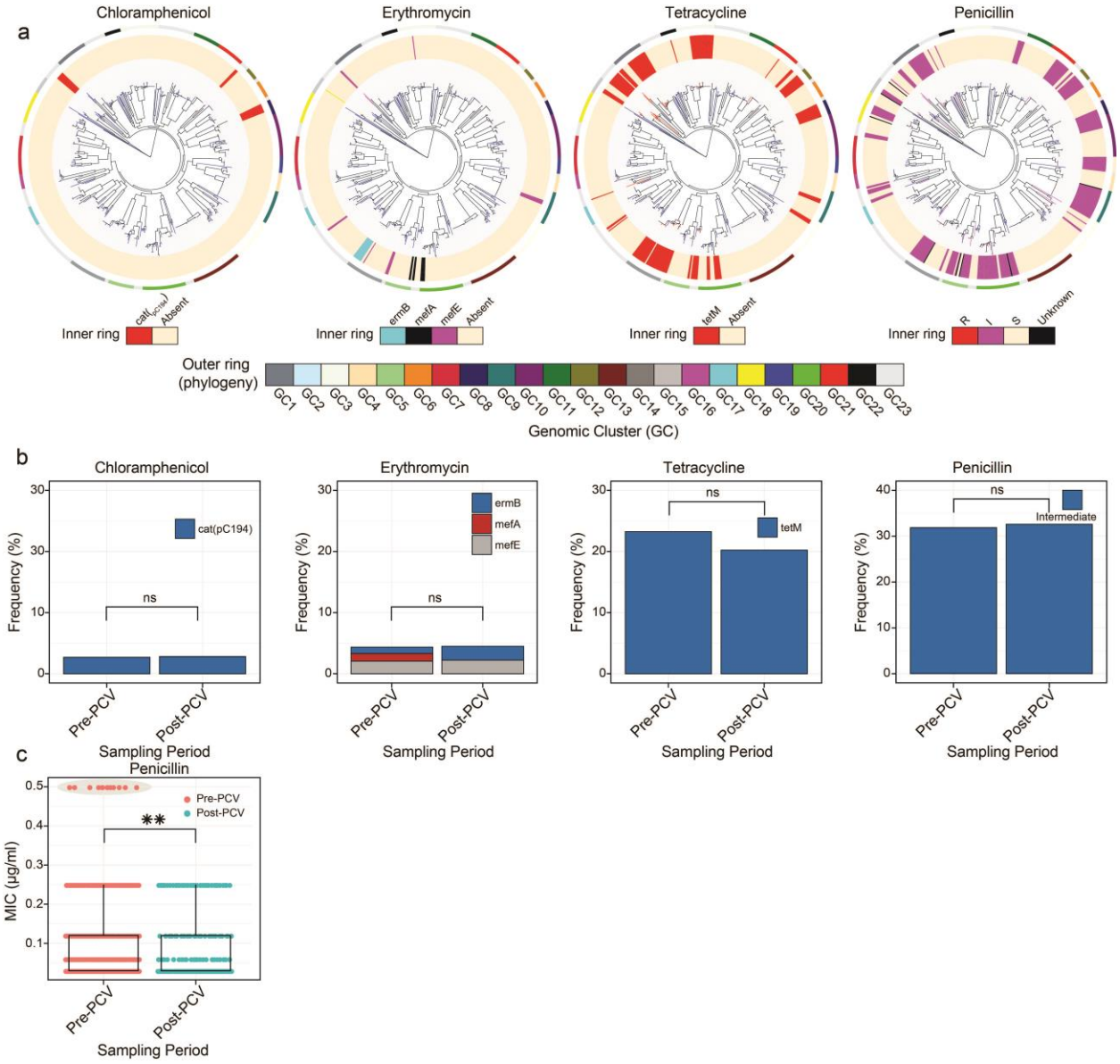
Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 5



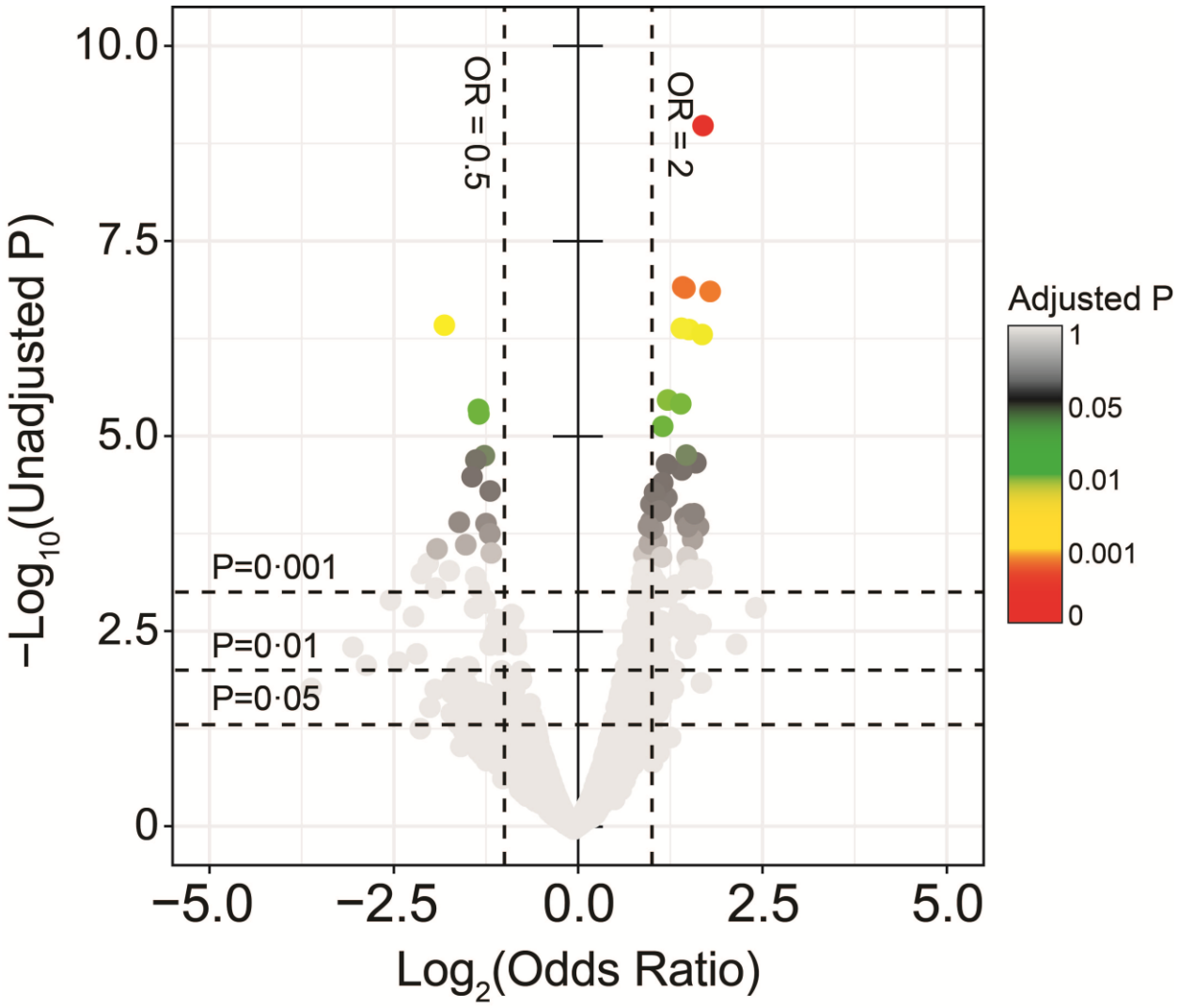
Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 6



Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019

Figure 7



Downloaded from https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciz404/5490137 by The University of Edinburgh user on 19 June 2019